

MULTIMEDIA DATA TRANSMISSION SYSTEM

DESCRIPTION

Technical field

This invention relates to a multimedia data transmission system.

State of prior art

5 Conventional multimedia servers are designed to be accommodated on a single platform. Usually, they consist simply of an application that runs on a computer equipped with interface cards to the telephone network.

10 In its most widely distributed form, a host server is capable of finding data on external data servers accessible through the same LAN, using RPC (Remote Procedure Call) or ODBC (Open DataBase Connectivity) type protocols.

15 This type of structure is suitable for the accommodation of simple multimedia servers in which there is no dynamic information. A company that would like to have a server accommodated describes the required service logic (if the user types #1, "you
20 typed 1"... should be displayed) statically, and this logic runs on the service supplier accommodation platform independently.

 On the other hand, it becomes impossible to accommodate an application that requests information
25 that necessitates close integration with one of the company's vital databases (booking statements, etc.), and the company must equip itself with its own infrastructure.

 More and more companies would like to integrate
30 this type of multimedia service more closely with

internal data in their industrial process. The objective is to inform the customer in real time if the ticket that he has just purchased is available, the value of his share portfolio, etc. These are dynamic data that are only available within the company.

Conventional multimedia accommodation services are not capable of satisfying these requirements, such that requesting companies are obliged to install their own server with the associated investments (private telephone exchange, telephone lines, etc.).

In order to overcome the disadvantages of this type of server, the invention proposes a multimedia data transmission system, the purpose of which is to provide a dynamic multimedia service for companies who would like it, without obliging the company to purchase any hardware and while making a server accessible to the company using several technologies (particularly from the telephone network and from the Internet network), with fully transparent service logic.

Description of the invention

The system according to the invention relates to a multimedia data transmission system characterized in that it comprises a WAN, in which the confidentiality and security are not controlled from end to end, onto which a shared voice and/or video resources host server designed to provide a dynamic service to at least one user, and at least one call control server located at each service supplier are connected.

Advantageously, the host server connected to the network through an interface is composed of five subsystems:

- A protocol stack subsystem with an interface that:
 - receives calls from the data network at the exchange;

- 5
SUB A2
54B A3
54B A4
54B A5
15
20
25
- detects incoming calls and captures caller and called party numbers;
 - detects dial tones;
 - generates arbitrary coding-decoding media data streams;
 - receives arbitrary media coding-decoding data streams.
 - A command interpreter subsystem capable of:
 - generating messages on detection of new calls to a call control server placed at a customer;
 - generating event messages;
 - making use of commands originating from call control servers placed at customers, such as:
 - * order to play a pre-recorded audio or video file,
 - * order to synthesize a voice message starting from a text,
 - * order to start waiting for a dial tone,
 - * order to disconnect the call,
 - * order for voice recognition or other application.
 - A high performance transcoding resource subsystem.
 - A voice synthesis and/or video resource subsystem.
 - An audio or video sequences recording/reproduction module subsystem.

Advantageously, ^{where in fig?} each call control server located at a customer is software that receives events signaled by the host server and sends commands in reaction to these events. This software can run on a computer equipped with two network interfaces, one connected to the WAN to communicate with the host server, and the other connected to a company private network in order to dialog with databases and other industrial processes belonging to the customer.

Thus, a new generation "accommodation" service can be provided in which all expensive resources (voice synthesis cards, etc.) are shared, while the customer

maintains control over the ^{? what application} application and can interface it with whatever resources he wishes.

Brief description of the drawings

- 5 - Figure 1 illustrates a first embodiment of the invention;
- 34B
A7 } - Figure 2 illustrates the dialog between an operator server with voice recognition and the server belonging to a company A;
- 10 - Figure 3 illustrates an example of a voice recognition procedure;
- Figure 4 illustrates an embodiment of a specialized page that reacts to voice.

Detailed presentation of an embodiment

15 The invention relates to a multimedia data transmission system that comprises a WAN, which may or may not be public, on which the confidentiality and security are not controlled from end to end, and onto

54B
A8 } 20 which a shared voice and/or video resources host server is connected and provides a dynamic service to at least one customer, and onto which at least one call control server located at each customer is also connected.

 The invention consists of placing a voice resource

25 in the WAN (capable of reproducing audio files, recording them, performing synthesis or voice recognition, detecting DTMF (Dual Tone MultiFrequency) tones from two sounds, equipped with a protected protocol that can remote control it from a wide area

30 network (such as the Internet network).

 The application that controls this voice resource may be located anywhere on the network. Thus, the

server is a distributed platform in which expensive resources are located in the network, and in which the

voice resource served ?
or
the application served ?

service logic (software only) is located at the customer.

5 Therefore, the invention can be used to share the voice resource server located in the network of an operator between several customers that execute the service logic in their premises. The companies simply need to have a connection with the data network. The operator server is accessible either from multimedia stations connected to the data network, or from any
10 telephone through a gateway.

With the invention, the supplier of the "accommodation" service provides a call control software to his customers, who run it locally on a machine in their network, and interface it with their
15 critical databases.

When a call arrives for this customer, it reaches the shared voice resource platform. This platform analyzes the requested number or the "ALIAS" for IP (INTERNET PROTOCOL) calls and deduces the client concerned. It sends a new call notification through the WAN to the call control application for the customer concerned. In particular, this application may ask the following in return:

- play a prerecorded audio file;
- 25 - synthesize a text;
- record a text;
- ask for a video sequence to be sent if the connected person has an appropriate terminal;
- make a voice recognition.

30 The voice resource can be made above the H.323 protocol so that users can be connected through the switched telephone network (through an STN/IP gateway), or through the Internet network, indifferently.

35

In one advantageous embodiment, the ⁷host server is connected to the WAN through an Ethernet or other interface, and is composed of five subsystems:

- A first subsystem, which is an H.323 protocol stack,
 5 for which the API (Application Programming Interface) is capable of:
 - detecting incoming calls and capturing the caller and called party numbers (or H.323 ALIAS);
 - detecting DTMF tones (transported in the H.245 protocol);
 - 10 - generating media data streams (sound + video) with arbitrary coding-decoding;
 - receiving media data streams (sound + video) with arbitrary coding-decoding;
- 15 • Possibly a second subsystem, which is a high performance transcoding resource, typically a digital signal processor card capable of transcoding the G.711 / G.723.1 protocols.
- 20 • Possibly a third subsystem which is a voice synthesis resource generating G.711 or G.723.1 type data streams, possibly with "streaming" capacities (division of a large file into successive small elements with limited duration).
- 25 • Possibly a fourth subsystem, which is an audio and video sequence recording / reproduction module with "streaming" functions during reproduction.

30

The action of these subsystems is coordinated by a fifth subsystem which is essentially a command interpreter capable of:

- generating new call detection messages to a call control server placed at a customer; it must

35

54B
A14

54B
A15

- also choose the right call control server starting from the called number;
- generating event messages, for example corresponding to DTMF tones;
 - implementing commands from call control servers placed at customers, such as:
 - * order to play a prerecorded audio or video file,
 - * order to synthesize a voice message from a text,
 - * order to go in waiting for a DTMF dial tone,
 - * order to disconnect the call,
 - * order for voice recognition or other application.

10

15

Calls from the switched telephone network are translated by an STN network/H.232 gateway for processing by the host server. The gateway function may possibly be integrated in the host server.

20

Other subsystems (voice recognition, fax generation/reception, etc.) may be added to increase the functional richness of the complete assembly.

54B
A16

25

In one advantageous embodiment, the call control server located at the customer is a simple software (for example "Window NT" service) that receives events signaled by the host server and sends commands in reaction to these events. This software may run on a computer provided with two network interfaces, one connected to the Internet network to communicate with the host server, and the other connected to a company private network to dialog with databases and other industrial processes within the company.

30

The host computer is configured so as to not transmit IP packets from the Internet network to the internal network.

The customer can configure the service logic itself
 5 using a script language (for example Java Script, VisualBasic), or a graphic interface.

The dialog protocol may be any secure dialog protocol with short waiting times. In one embodiment,
 10 a protocol is used on a standard UDP in which each information block sent is in the following form:

```
<block><random><64 random bits></random><cipherblock>
                                encrypted data</cipherblock> </block>
```

15 The encrypted information block must have the following structure once it has been decrypted:

```
<clearinfo>
20 <serial>serial number</serial>
   <other information> ... <other information>
</clearinfo>
```

Information encrypted in the "cipherblock" block is
 25 obtained by encrypting the "clearinfo" structure using the DES (Data Encryption Standard) standard in CBC (Cipher Block Chaining) mode, using the 64 random bits for the initial exclusive OR. The sender's identity is proven by the possibility of finding an intelligible
 30 message with decryption. The receiver must memorize the last serial number received from the sender and discard any message received with a serial number less than or equal to the current serial number.

The sender can protect his transmission (UDP
 35 standard) by sending several identical messages. The

receiver memorizes the serial number of the first correctly received message and discards subsequent messages without examining them.

5 Figure 1 illustrates a first example use, which is for the communication by an IP interactive voice server.

10 A WAN network 10, for example Internet, in which the voice and/or video resource operator server 11 is connected to:

- an ordinary telephone 12 through a WAN telephone gateway 13;
- a multimedia station 14 through a two-directional link 15, of the H.323, SIP, or other type of voice data stream;
- three servers 16, 17 and 18 for companies A, B and C.

15 When the operator server 11 receives a new communication from a user, the first thing it does is to analyze the called number and then deduces which company server should manage the communication; for example server 16 for company A.

20 Company A makes fast part orders. Server 16 sends its welcome announcement stored in the welcome file in the operator server 11: "welcome to company A's fast order server, please press on the '*' key to begin". Informed users can interrupt this announcement by pressing on the '*' key.

25 As soon as the user presses on '*', the operator server 11 informs company A's server 16 with a "DTMF event" message. Company A's server 16 then begins to play the "Do_you_want_to_order" file which contains a recording of this phrase.

30 Company A's server 16 decides to use the voice command, to order the operator server 11 to start

recognition on the "yes, no" vocabulary. As soon as the user says "yes", the server 16 is informed by a "Word_recognition" message.

Server 16 then asks how many parts the customer wants to order and records this number by voice recognition. It then stops the voice recognition procedure by a "Stop_recognition" command.

Finally, the server 16 repeats the amount of the order to the customer asking the operator server 11 to synthesize the "You have ordered three parts" character string. The user then hangs up.

The dialog between the operator server 11 with voice recognition which receives an H.323, SIP or other voice data stream and company A's server 16, is illustrated in figure 2.

Voice recognition procedures usually comprise two parts:

- the first part (A) uses the voice data stream (64 kbits for standard G.711 and 6.4 kbits for standard G.723.1) and extracts significant components from it (spectrum, etc.), the result is a low rate data stream between 4 and 8 kbits/s;
- the second part (B) attempts to recognize words in a vocabulary starting from components transmitted by the first part A.

The scheme illustrated in figure 3 shows how the different modules of a voice recognition procedure communicate with each other.

There are two ways of creating a voice recognition procedure in the IP interactive server:

- When the customer who is calling the company server is not controlled by the network operator, the A and B components have to be put on the

operator server. This is the method used in the above example.

- However if the network operator can, it is better to extract significant components at the customer in order to make less use of the passband on the network between the customer and the operator server. This extraction phase requires very little calculation power.

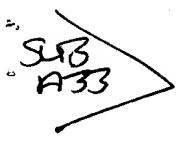
For example, if the client is an IP telephony software, the significant components extraction module may appear like a new speech encoder. The operator server then negotiates with the customer for use of this encoder during the connection.

Another possible embodiment is to put a software component in a specialized displayed HTML page (ActiveX or Java) that interfaces with voice resources on the customer station and only sends significant components of the voice data stream to the operator server. Thus, a specialized page can be created which reacts to voice, as in the example in figure 3.

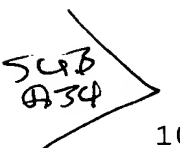
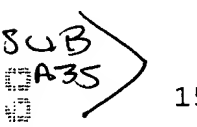
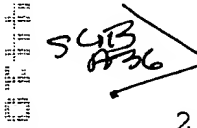
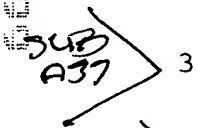
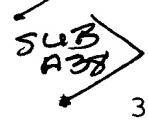
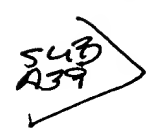
Figure 4 illustrates another possible example embodiment with the IP audiotel server, for a specialized page that reacts to voice.

In this example embodiment, the customer is a software object ("ActiveX or Java") integrated in a specialized page. This object sends significant voice data stream components input on the customer station computer to the operator server. It can do this using the FTP protocol on the IP network, or simply the TCP protocol if the reaction time is not a major constraint.

The operator server recognizes words in this data stream and informs the company server of recognized words.

5  The company server then initiates actions as a function of the recognized words. For example, it can send a command message to the ActiveX component to display another specialized page.

The following protocol is used:

- 10  1. Connection request: Connection request message (operator server => company server)
(Implicit in TCP/IP by opening the exchange mechanism in TCP/IP)
- 15  2. Call data: Transmit call data (operator server => company server)
Called number
Calling number
- 20  3. Read sound: Read a sound file (company server => operator server)
Logical channel number
Name of the element to which the response is to be notified
Time in ms before playing the sound
File name to be played
Digit used to detect the end of the sound file
25 Format of the sound file (Wav, Vox, ADPCM ...)
Data format
Sampling frequency
- 30  4. DTMF event message (operator server => company server)
Logical channel number
DTMF key code
- 35  5. Sound recording: Recording of a message (company server => operator server)
Channel number
Name of the element to which the response is to be notified
Time before beginning the recording
Name of the message save file
End of recording character
40 Maximum recording time
Maximum silence time
Save file format
Data format
Sampling frequency
45 Send a beep to signal when the recording starts
- 50  6. Send tone: Send a tone (company server => operator server)
Channel
Name of the element to which the response is to be notified
TimeBefore

Dial Tone
 Frequency 1
 Frequency 2
 Amplitude 1
 Amplitude 2
 Tone duration

5

7. Read chain: Concatenate a string of characters (company server => operator server)

10

Logical channel number
 Name of the element to which the response is to be notified
 Time before reading sound
 Character string, for which the data => sound conversion is to be made

15

End of file character string
 Sound file format (Wav, Vox, ADPCM ...)
 Data format

Sampling frequency format

20

Mix size, so that two files can be mixed later (Smooth transition)

Breakdown type, which will be used later for number generation time functions starting from a sound library
 Character used to separate expressions in the character string

25

File name resulting from the concatenation
 Word field name
 Sound field name
 Dictionary access path

30

8. Disconnect user: The caller hung up (operator server => Company server)

Logical channel number to be disconnected
 (Implicit in TCP/IP by closing the TCP/IP exchange mechanism)

35

9. Disconnect server: Disconnection request by the company server software (company server => operator server)

Logical channel number to be disconnected

40

10. Voice synthesis:

Logical channel number
 Name of the element to which the response is to be notified
 Text to be converted in voice synthesis
 Choose a specific voice, if required
 Speaking speed
 Speaking frequency

45

...

11. Extended call (function of the call transfer request)

50

Logical channel number
 Name of the element to which the response is to be notified
 Transfer request time
 Number to which the call is to be transferred
 Call type

Number of rings before abandon
Time to analyze the result of the transfer request

- 5 12. Start recognition (function requesting beginning of voice recognition)
Logical channel number
Name of the element to which the response is to be notified
Name of the words file to be analyzed
Digit used to detect the end of the sound file
10 Maximum recording time
Maximum silence time
Send a "beep" signaling the beginning of the recording
- 15 13. Stop recognition (function requesting the beginning of voice recognition)
Logical channel number
- 20 14. Word recognition (function requesting the beginning of voice recognition)
Logical channel number
Name of the element to which the response is to be notified
List of recognized words

25 We will now describe several other example embodiments.

• *Call from the telephone network*

A person who would like to book a journey calls 0836011234. This number actually connects to an
30 STN/H.323 network gateway that converts the call into IP data and sends it to the host voice resources server.

35 The voice resources server analyzes the requested number and deduces that the call must be controlled by the call control server located at the IP address 192.12.13.14 (located in the travel agent).

40 Therefore, it sends a new call message to the travel agent's call control server. This call control server asks it to play a musical background quickly presenting the company and asking the caller to press "1" to book a voyage, or "2" to leave a message.

543
A45

The person presses "1" and the host voice resources server/retransmits the event to the travel agent's call control server.

543
A46

5 The dialog continues. It could be imagined that the travel agent would like to announce the price of a particular voyage. The call control server looks in the travel agent's database for prices and availabilities, and asks the host voice resources server to play the recorded string "the price of your
10 voyage is", and then to synthesize "2345" and then play "Francs".

• *Call from the Internet network*

15 An H.323 terminal clicks on a link starting from a travel agent's Internet site, provoking a call from the H.323 terminal to the H.323 host server. The server analyzes the called number and sends an indication for the new call to the travel agent's call control server.

20 The travel agent's call control server does not need to be modified, and can execute the same scenario as in the previous case.

25 But it can also choose to offer more services, since a protocol element informs it at the time of the indication of the new call that the call is incoming from the Internet network, it can suggest that a specific page should be viewed, or even give the order to the host server to play a video sequence describing a particular voyage.

30 The call is free for the Internet network user.

• *Call from another country*

543
A48

35 If the operator has installed another host voice resources server in another country, the travel agent may be accessible from this country. The operator simply reserves a number that is forwarded to the local

543
A45

voice resources server. The server continues to contact the company's call control server. The source of the call is indicated when a new call indication is received, so that the call control server can dynamically adapt to the most suitable language when it is helpful to do so.

This solution is much less expensive than a conventional solution, since no international voice communication is necessary.